Homework Assignment 3
601.467 / 667 Introduction to Human Language Technology
Fall 2019
Due: Wednesday November 20, 2019

November 5, 2019

Sequence-to-Sequence Modeling

In this assignment, you will implement two sequence to sequence (seq2seq) models that predict the pronunciation of an English word.

Note1: This is a long assignment, start early.

Note2: You may work in groups of (max) 3 people.

Question 1. Open the following colab notebook and complete the assignment:

https://colab.research.google.com/drive/1wKZg6IwaWE1urXpU0ZP1toKOzU5mcmWF

Follow all the steps specified in it. Include link to your solved notebook in your submission. There are detailed `TODOs` specified in the colab link, you should complete all of them.

Report the following:

1. The number of params used by the simple (non-Attention) Seq2Seq model.

2. Training and Validation loss for the simple (non-Attention) Seq2Seq Model. Our implementation attains $0.6244, 0.6932$ for training and validation loss at the end of 15 epochs.

3. Training and Validation accuracies for the simple (non-Attention) Seq2Seq Model. Our implementation attains $0.80\%, 0.79\%$ training and validation accuracy at the end of 15 epochs.

4. The number of params used by the Attention Seq2Seq model.

5. Training and Validation loss for the Attention Seq2Seq Model. Our implementation attains $0.3913, 0.5606$ for training and validation loss at the end of 10 epochs.

6. Training and Validation accuracies for the Attention Seq2Seq Model. Our model attains $0.87\%, 0.84\%$ training and validation accuracy at the end of 10 epochs.(you can paste these into the pdf from the output log)

7. Test Avg. CER for the Simple Seq2Seq model

8. Test Avg. CER for the Attention Seq2Seq model

Answer the following:

1. Do you notice any pattern in the improvements made by the Attention Seq2Seq model over the Simple Seq2Seq model? Take a look at the output of the evaluation method for the two models.

2. Look at the forward function of the EncoderDecoder class. What is the purpose of the following lines:

```
y_input  = y[:,  :-1]
y_output = y[:,  1:]
```

3. In previous homework the generation function had a temperature term. What implicit temperature are we using this homework?

Optional Part: Try to improve on the Attention based model you have implemented! You are free to increase the number of parameters or try one of these attention mechanisms:

1. A Simple dot product attention:

$$\alpha_{i,j} = H \cdot D \tag{1}$$

2. Attention weights computed from a feed forward network:

$$\alpha_{i,j} = \tanh(f([H; D])) \tag{2}$$